

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

- 5 (i) APPLICANT: Deen, Keith C.  
Dillon, Susan B.  
Porter, Terence C.  
Sweet, Raymond A.
- 10 (ii) TITLE OF THE INVENTION: Human Monoclonal Antibodies
- (iii) NUMBER OF SEQUENCES: 30
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: SmithKline Beecham Corporation  
(B) STREET: 709 Swedeland Road  
(C) CITY: King of Prussia  
(D) STATE: PA  
(E) COUNTRY: U.S.A.
- 20 (F) ZIP: 19046
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible
- 25 (C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 30 (B) FILING DATE:  
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: 60/030,149
- 35 (B) FILING DATE: 01-NOV-1997

## (viii) ATTORNEY/AGENT INFORMATION:

- 5 (A) NAME: Geiger, Kathleen  
(B) REGISTRATION NUMBER: 35,880  
(C) REFERENCE/DOCKET NUMBER: P50504

## (ix) TELECOMMUNICATION INFORMATION:

- 10 (A) TELEPHONE: 610-270-5968  
(B) TELEFAX: 610-270-5090  
(C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
20 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

25

Ser Ile Thr Gly Gly Ser Asn Gly Ile Asn Tyr Ala Asp Ser Val Lys

1

5

10

15

Arg

30

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids  
35 (B) TYPE: amino acid  
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ser Ile Thr Gly Gly Ser Asn Gly Ile Asn Tyr Ser Asp Ser Val Lys  
1 5 10 15  
Arg

10

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ser Ile Thr Gly Gly Ser Asn Gly Ile Gln Tyr Ser Asp Ser Val Lys  
25 1 5 10 15  
Arg

30 (2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 98 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

5    Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly  
      1                                5                                10                                15  
      Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr  
                              20                                25                                30  
      Glu Met Asn Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val  
 10                                35                                40                                45  
      Ser Tyr Ile Ser Ser Ser Gly Ser Thr Ile Tyr Tyr Ala Asp Ser Val  
                              50                                55                                60  
      Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr  
      65                                70                                75                                80  
 15    Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys  
                              85                                90                                95  
      Ala Arg

20                                (2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 139 amino acids

(B) TYPE: amino acid

25                                (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

30                                (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly  
      1                                5                                10                                15  
      Val His Ser Glu Val Gln Leu Leu Glu Val Glu Ser Gly Gly Gly Leu  
 35                                20                                25                                30

Arg Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Thr Thr  
 35 40 45  
 Leu Ser Gly Tyr Thr Met His Trp Val Arg Gln Ala Pro Gly Lys Gly  
 50 55 60  
 5 Leu Glu Trp Val Ser Ser Ile Thr Gly Gly Ser Asn Phe Ile Asn Tyr  
 65 70 75 80  
 Ser Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys  
 85 90 95  
 Asn Ser Leu Tyr Leu Gln Met Asn Ser Leu Thr Ala Glu Asp Thr Ala  
 10 100 105 110  
 Val Tyr Tyr Cys Ala Thr Ala Pro Ile Ala Pro Pro Tyr Phe Asp His  
 115 120 125  
 Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser  
 130 135

15

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 138 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Leu Leu Arg Gly  
 1 5 10 15  
 30 Val Gln Cys Gln Val Gln Leu Val Val Glu Ser Gly Gly Gly Leu Arg  
 20 25 30  
 Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Thr Thr Leu  
 35 40 45  
 Ser Gly Tyr Thr Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu  
 35 50 55 60

Glu Trp Val Ser Ser Ile Thr Gly Gly Ser Asn Phe Ile Asn Tyr Ser  
 65                      70                      75                      80  
 Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn  
                     85                      90                      95  
 5 Ser Leu Tyr Leu Gln Met Asn Ser Leu Thr Ala Glu Asp Thr Ala Val  
                     100                      105                      110  
 Tyr Tyr Cys Ala Thr Ala Pro Ile Ala Pro Pro Tyr Phe Asp His Trp  
                     115                      120                      125  
 Gly Gln Gly Thr Leu Val Thr Val Ser Ser  
 10                      130                      135

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 138 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 20 (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Leu Leu Arg Gly  
 25    1                      5                      10                      15  
 Val Gln Cys Gln Val Gln Leu Val Val Glu Ser Gly Gly Gly Leu Arg  
                     20                      25                      30  
 Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Thr Thr Leu  
                     35                      40                      45  
 30 Ser Gly Tyr Thr Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu  
                     50                      55                      60  
 Glu Trp Val Ser Ser Ile Thr Gly Gly Ser Asn Phe Ile Asn Tyr Ala  
 65                      70                      75                      80  
 Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn  
 35                      85                      90                      95

Ser Leu Tyr Leu Gln Met Asn Ser Leu Thr Ala Glu Asp Thr Ala Val  
                   100                          105                          110  
 Tyr Tyr Cys Ala Thr Ala Pro Ile Ala Pro Pro Tyr Phe Asp His Trp  
                   115                          120                          125  
 5 Gly Gln Gly Thr Leu Val Thr Val Ser Ser  
                   130                          135

## (2) INFORMATION FOR SEQ ID NO:8:

- 10 (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 138 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: linear

- 15 (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

20 Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Leu Leu Arg Gly  
    1                          5                          10                          15  
 Val Gln Cys Gln Val Gln Leu Val Val Glu Ser Gly Gly Gly Leu Arg  
                   20                          25                          30  
 Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Thr Thr Leu  
 25                  35                          40                          45  
 Ser Gly Tyr Thr Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu  
                   50                          55                          60  
 Glu Trp Val Ser Ser Ile Thr Gly Gly Ser Asn Phe Ile Gln Tyr Ser  
                   65                          70                          75                          80  
 30 Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn  
                   85                          90                          95  
 Ser Leu Tyr Leu Gln Met Asn Ser Leu Thr Ala Glu Asp Thr Ala Val  
                   100                          105                          110  
 Tyr Tyr Cys Ala Thr Ala Pro Ile Ala Pro Pro Tyr Phe Asp His Trp  
 35                  115                          120                          125

Gly Gln Gly Thr Leu Val Thr Val Ser Ser  
130 135

## (2) INFORMATION FOR SEQ ID NO:9:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

10

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

15

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly  
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Ser Ile Ser Ser Tyr  
20 25 30

20

Leu Asn Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile  
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly  
50 55 60

25

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro  
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys  
85

## (2) INFORMATION FOR SEQ ID NO:10:

30

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 124 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

35

(D) TOPOLOGY: linear



(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

5  Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly
   1             5             10             15
Val His Ser Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val
   20             25             30
Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Thr Gln Ser Val Ser Asn
10             35             40             45
Phe Leu Asn Trp Tyr Gln Gln Lys Pro Gly Glu Ala Pro Thr Leu Leu
   50             55             60
Ile Tyr Asp Ala Ser Thr Ser Gln Ser Gly Val Pro Ser Arg Phe Ser
   65             70             75             80
15 Gly Ser Gly Ser Gly Met Asp Phe Ser Leu Thr Ile Ser Ser Leu Gln
   85             90             95
Pro Glu Asp Leu Ala Met Tyr Tyr Cys Gln Ala Ser Ile Asn Thr Pro
   100            105            110
Leu Phe Gly Gly Gly Thr Arg Ile Asp Met Arg Arg
20             115            120

```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

```

25  (A) LENGTH: 101 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

```

30 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp Leu Arg
35  1             5             10             15

```

Gly Ala Arg Cys Asp Ile Gln Met Asn Phe Leu Asn Trp Tyr Gln Gln  
                   20                          25                          30  
 Lys Pro Gly Glu Ala Pro Thr Leu Leu Ile Tyr Asp Ala Ser Thr Ser  
                   35                          40                          45  
 5 Gln Ser Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Met Asp  
                   50                          55                          60  
 Phe Ser Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Leu Ala Met Tyr  
                   65                          70                          75                          80  
 Tyr Cys Gln Ala Ser Ile Asn Thr Pro Leu Phe Gly Gly Gly Thr Arg  
 10                                  85                                  90                                  95  
 Ile Asp Met Arg Arg  
                                   100

## (2) INFORMATION FOR SEQ ID NO:12:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 106 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

20

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

25

Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln  
   1                          5                          10                          15  
 Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr  
                   20                          25                          30  
 30 Pro Arg Glu Ala Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser  
                   35                          40                          45  
 Gly Asn Ser Gln Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr  
                   50                          55                          60  
 Tyr Ser Leu Ser Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys  
 35                                  65                                  70                                  75                                  80

His Lys Val Tyr Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro  
                             85                            90                            95  
 Val Thr Lys Ser Phe Asn Arg Gly Glu Cys  
                             100                            105

5

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 106 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln  
   1                            5                            10                            15  
 20 Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr  
                             20                            25                            30  
 Pro Arg Glu Ala Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser  
                             35                            40                            45  
 Gly Asn Ser Gln Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr  
 25            50                            55                            60  
 Tyr Ser Leu Ser Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys  
   65                            70                            75                            80  
 His Lys Val Tyr Ala Cys Glu Val Thr His Gln Gly Leu Ser Leu Pro  
                             85                            90                            95  
 30 Val Thr Lys Ser Phe Asn Arg Gly Glu Cys  
                             100                            105

(2) INFORMATION FOR SEQ ID NO:14:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6284 base pairs

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

100555-555555

	GACGTCGCGG	CCGCTCTAGG	CCTCCAAAAA	AGCCTCCTCA	CTACTTCTGG	AATAGCTCAG	60
	AGGCCGAGGC	GGCCTCGGCC	TCTGCATAAA	TAAAAAAAT	TAGTCAGCCA	TGCATGGGGC	120
10	GGAGAATGGG	CGGAAGTGGG	CGGAGTTAGG	GGCGGGATGG	GCGGAGTTAG	GGGCGGGACT	180
	ATGGTTGCTG	ACTAATTGAG	ATGCATGCTT	TGCATACTTC	TGCCTGCTGG	GGAGCCTGGG	240
	GACTTTCCAC	ACCTGGTTGC	TGACTAATTG	AGATGCATGC	TTTGCACTACT	TCTGCCTGCT	300
	GGGGAGCCTG	GGGACTTTCC	ACACCCTAAC	TGACACACAT	TCCACAGAAT	TAATTCCCGG	360
	GGATCGATCC	GTCGACGTAC	GACTAGTTAT	TAATAGTAAT	CAATTACGGG	GTCATTAGTT	420
15	CATAGCCCAT	ATATGGAGTT	CCGCGTTACA	TAACTTACGG	TAAATGGCCC	GCCTGGCTGA	480
	CCGCCCAACG	ACCCCGCCCC	ATTGACGTCA	ATAATGACGT	ATGTTCCCAT	AGTAACGCCA	540
	ATAGGGACTT	TCCATTGACG	TCAATGGGTG	GACTATTTAC	GGTAAACTGC	CCACTTGGCA	600
	GTACATCAAG	TGTATCATAT	GCCAAGTACG	CCCCCTATTG	ACGTCAATGA	CGGTAAATGG	660
	CCCGCCTGGC	ATTATGCCCA	GTACATGACC	TTATGGGACT	TTCCTACTTG	GCAGTACATC	720
20	TACGTATTAG	TCATCGCTAT	TACCATGGTG	ATGCGGTTTT	GGCAGTACAT	CAATGGGCGT	780
	GGATAGCGGT	TTGACTCAGG	GGGATTTCCA	AGTCTCCACC	CCATTGACGT	CAATGGGAGT	840
	TTGTTTTGGC	ACCAAAATCA	ACGGGACTTT	CCAAAATGTC	GTAACAACTC	CGCCCCATTG	900
	ACGCAAATGG	GCGGTAGGCG	TGTACGGTGG	GAGGTCATA	TAAGCAGAGC	TGGGTACGTG	960
	AACCGTCAGA	TGCCTGGAG	ACGCCATCGA	ATTCTGAGCA	CACAGGACCT	CACCATGGGA	1020
25	TGGAGCTGTA	TCATCCTCTT	CTTGGTAGCA	ACAGCTACAG	GTGTCCACTC	CGAGGTCCAA	1080
	CTGCTCGAGG	AGTCTGGGGG	AGGCCTGGTC	AGGCCTGGCG	GGTCCCTAAG	ACTCTCGTGT	1140
	GCAGCCTCTG	GAACCAACCT	CAGTGGCTAT	ACCATGCACT	GGGTCCGCCA	GGCTCCAGGG	1200
	AAGGGGCTGG	AGTGGGTCTC	ATCCATTACT	GGAGGTAGCA	ACTTCATAAA	CTACTCAGAC	1260
	TCAGTGAAGG	GCCGATTAC	CATCTCCAGA	GACAACGCCA	AGAACTCACT	TTATCTGCAA	1320
30	ATGAACAGCC	TGACAGCCGA	GGACACGGCT	GTCTATTATT	GTGCGACCGC	CCCTATAGCA	1380
	CCGCCCTACT	TTGACCACTG	GGGCCAGGGA	ACCCTGGTCA	CCGTCTCCTC	AGCCTCCACC	1440
	AAGGGCCCAT	CGGTCTTCCC	CCTGGCACCC	TCCTCCAAGA	GCACCTCTGG	GGGCACAGCG	1500
	GCCCTGGGCT	GCCTGGTCAA	GGACTACTTC	CCCGAACCAG	TGACCGTGTC	GTGGAATCA	1560
	GGCGCCCTGA	CCAGCGGCGT	GCACACCTTC	CCGGCTGTCC	TACAGTCCTC	AGGACTCTAC	1620
35	TCCCTCAGCA	GCGTGGTGAC	TGTGCCCTCC	AGCAGCTTGG	GCACCCAGAC	CTACATCTGC	1680
	AACGTGAATC	ACAAGCCCAG	CAACACCAAG	GTGGACAAGA	AAGTTGAGCC	CAAATCTTGT	1740

	GACAAAATC	ACACATGCCC	ACCGTGCCCA	GCACCTGAAC	TCCTGGGGGG	ACCGTCAGTC	1800
	TTCTCTTCC	CCCCAAAACC	CAAGGACACC	CTCATGATCT	CCCGGACCCC	TGAGGTCACA	1860
	TGCGTGGTGG	TGGACGTGAG	CCACGAAGAC	CCTGAGGTCA	AGTTCAACTG	GTACGTGGAC	1920
	GGCGTGGAGG	TGCATAATGC	CAAGACAAAG	CCGCGGGAGG	AGCAGTACAA	CAGCACGTAC	1980
5	CGGGTGGTCA	GCGTCCCTAC	CGTCTGCAC	CAGGACTGGC	TGAATGGCAA	GGAGTACAAG	2040
	TGCAAGGTCT	CCAACAAAGC	CCTCCCAGCC	CCCATCGAGA	AAACCATCTC	CAAAGCCAAA	2100
	GGGCAGCCCC	GAGAACCACA	GGTGTAACAC	CTGCCCCCAT	CCCGGGATGA	GCTGACCAAG	2160
	AACCAGGTCA	GCCTGACCTG	CCTGGTCAAA	GGCTTCTATC	CCAGCGACAT	CGCCGTGGAG	2220
	TGGGAGAGCA	ATGGGCAGCC	GGAGAACAAC	TACAAGACCA	CGCCTCCCGT	GCTGGACTCC	2280
10	GACGGCTCCT	TCTTCTCTA	CAGCAAGCTC	ACCGTGGACA	AGAGCAGGTG	GCAGCAGGGG	2340
	AACGTCTTCT	CATGCTCCGT	GATGCATGAG	GCTCTGCACA	ACCACTACAC	GCAGAAGAGC	2400
	CTCTCCCTGT	CTCCGGGTAA	ATGATAGATA	TCTACGTATG	ATCAGCCTCG	ACTGTGCCTT	2460
	CTAGTTGCCA	GCCATCTGTT	GTTTGGCCCT	CCCCCGTGCC	TTCTTTGACC	CTGGAAGGTG	2520
	CCACTCCAC	TGCTCTTCC	TAATAAAATG	AGGAAATTGC	ATCGCATTGT	CTGAGTAGGT	2580
15	GTCATTCTAT	TCTGGGGGGT	GGGGTGGGGC	AGGACAGCAA	GGGGGAGGAT	TGGGAAGACA	2640
	ATAGCAGGCA	TGCTGGGGAT	GCGGTGGGCT	CTATGGAACC	AGCTGGGGCT	CGACAGCGCT	2700
	GGATCTCCCG	ATCCCCAGCT	TTGCTTCTCA	ATTTCTTATT	TGCATAATGA	GAAAAAAGG	2760
	AAAATTAATT	TTAACACCAA	TTCAGTAGTT	GATTGAGCAA	ATGCGTTGCC	AAAAAGGATG	2820
	CTTTAGAGAC	AGTGTCTCT	GCACAGATAA	GGACAAACAT	TATTCAGAGG	GAGTACCCAG	2880
20	AGCTGAGACT	CCTAAGCCAG	TGAGTGGCAC	AGCATTCTAG	GGAGAAATAT	GCTTGTCATC	2940
	ACCGAAGCCT	GATTCCGTAG	AGCCACACCT	TGGTAAGGGC	CAATCTGCTC	ACACAGGATA	3000
	GAGAGGGCAG	GAGCCAGGGC	AGAGCATATA	AGGTGAGGTA	GGATCAGTTG	CTCCTCACAT	3060
	TTGCTTCTGA	CATAGTTGTG	TTGGGAGCTT	GGATAGCTTG	GACAGCTCAG	GGCTGCGATT	3120
	TCGCGCCAAA	CTTGACGGCA	ATCCTAGCGT	GAAGGCTGGT	AGGATTTTAT	CCCCGCTGCC	3180
25	ATCATGGTTC	GACCATTGAA	CTGCATGCTC	GCCGTGTCCC	AAAATATGGG	GATTGGCAAG	3240
	AACGGAGACC	TACCCTGGCC	TCCGCTCAGG	AACGAGTTCA	AGTACTTCCA	AAGAATGACC	3300
	ACAACCTCTT	CAGTGAAGG	TAAACAGAAT	CTGGTGATTA	TGGGTAGGAA	AACCTGGTTC	3360
	TCCATTCTCTG	AGAAGAATCG	ACCTTTAAAG	GACAGAATTA	ATATAGTTCT	CAGTAGAGAA	3420
	CTCAAAGAAC	CACCACGAGG	AGCTCATTTT	CTTGCCAAAA	GTTTGGATGA	TGCCTTAAGA	3480
30	CTTATTGAAC	AACCGGAATT	GGCAAGTAAA	GTAGACATGG	TTTGGATAGT	CGGAGGCAGT	3540
	TCTGTTTACC	AGGAAGCCAT	GAATCAACCA	GGCCACCTTA	GACTCTTTGT	GACAAGGATC	3600
	ATGCAGGAAT	TTGAAAGTGA	CACGTTTTTC	CCAGAAATTG	ATTTGGGGAA	ATATAAACTT	3660
	CTCCCAGAAT	ACCCAGGCGT	CCTCTCTGAG	GTCCAGGAGG	AAAAAGGCAT	CAAGTATAAG	3720
	TTTGAAGTCT	ACGAGAAGAA	AGACTAACAG	GAAGATGCTT	TCAAGTTCTC	TGCTCCCCTC	3780
35	CTAAAGCTAT	GCATTTTTAT	AAGACCATGG	GACTTTTGCT	GGCTTTAGAT	CAGCCTCGAC	3840
	TGTGCCTTCT	AGTTGCCAGC	CATCTGTTGT	TTGCCCCCTC	CCCGTGCCTT	CCTTGACCCT	3900

	GGAAGGTGCC	ACTCCCACTG	TCCTTTCCTA	ATAAAATGAG	GAAATTGCAT	CGCATTGTCT	3960
	GAGTAGGTGT	CATTCTATTC	TGGGGGGTGG	GGTGGGGCAG	GACAGCAAGG	GGGAGGATTG	4020
	GGAAGACAAT	AGCAGGCATG	CTGGGGATGC	GGTGGGCTCT	ATGGAACCAG	CTGGGGCTCG	4080
	ATCGAGTGTA	TGACTGCGGC	CGCGATCCCG	TCGAGAGCTT	GGCGTAATCA	TGGTCATAGC	4140
5	TGTTTCCTGT	GTGAAATGT	TATCCGCTCA	CAATTCCACA	CAACATACGA	GCCGGAAGCA	4200
	TAAAGTGTA	AGCCTGGGGT	GCCTAATGAG	TGAGCTAACT	CACATTAATT	GCGTTGCGCT	4260
	CACTGCCCCG	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCT	GCATTAATGA	ATCGGCCAAC	4320
	GCGCGGGGAG	AGGCGGTTTG	CGTATTGGGC	GCTCTTCCGC	TTCCTCGCTC	ACTGACTCGC	4380
	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA	CTCAAAGCGG	GTAATACGGT	4440
10	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG	AGCAAAAGGC	CAGCAAAAGG	4500
	CCAGGAACCG	TAAAAAGGCC	GCCTTGCTGG	CGTTTTTCCA	TAGGCTCCGC	CCCCCTGACG	4560
	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA	CCCACAGGA	CTATAAAGAT	4620
	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC	TGTTCCGACC	CTGCCGCTTA	4680
	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAAGCTGGC	GCTTTCTCAA	TGCTCACGCT	4740
15	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTT	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC	4800
	CCGTTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACATATC	TCTTGAGTCC	AACCCGGTAA	4860
	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG	GATTAGCAGA	GCGAGGTATG	4920
	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA	CGGCTACACT	AGAAGGACAG	4980
	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG	AAAAAGAGTT	GGTAGCTCTT	5040
20	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT	TGTTTGCAAG	CAGCAGATTA	5100
	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TTCTACGGGG	TCTGACGCTC	5160
	AGTGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG	ATTATCAAAA	AGGATCTTCA	5220
	CCTAGATCCT	TTTAAATTAA	AAATGAAGTT	TTAAATCAAT	CTAAAGTATA	TATGAGTAAA	5280
	CTTGGTCTGA	CAGTTACCAA	TGCTTAATCA	GTGAGGCACC	TATCTCAGCG	ATCTGTCTAT	5340
25	TTCGTTTCATC	CATAGTTGCC	TGACTCCCCTG	TCGTGTAGAT	AACTACGATA	CGGGAGGGCT	5400
	TACCATCTGG	CCCCAGTGCT	GCAATGATAC	CGCGAGACCC	ACGCTCACCG	GCTCCAGATT	5460
	TATCAGCAAT	AAACCAGCCA	GCCGGAAGGG	CCGAGCGCAG	AAGTGGTCCT	GCAACTTTAT	5520
	CCGCCTCCAT	CCAGTCTATT	AATTGTTGCT	GGGAAGCTAG	AGTAAGTAGT	TCGCCAGTTA	5580
	ATAGTTTGCG	CAACGTTGTT	GCCATTGCTA	CAGGCATCGT	GGTGTACACG	TCGTGCTTTG	5640
30	GTATGGCTTC	ATTGAGCTCC	GGTTCCCAAC	GATCAAGGCG	AGTTACATGA	TCCCCCATGT	5700
	TGTGCAAAAA	AGCGGTTAGC	TCCTTCGGTC	CTCCGATCGT	TGTCAGAAGT	AAGTTGGCCG	5760
	CAGTGTTATC	ACTCATGGTT	ATGGCAGCAC	TGCATAATTC	TCTTACTGTC	ATGCCATCCG	5820
	TAAGATGCTT	TTCTGTGACT	GGTGAGTACT	CAACCAAGTC	ATTCTGAGAA	TAGTGATATG	5880
	GGCGACCGAG	TTGCTCTTGC	CCGGCGTCAA	TACGGGATAA	TACCGCGCCA	CATAGCAGAA	5940
35	CTTTAAAGT	GCTCATCATT	GGAAAACGTT	CTTCGGGGCG	AAAACCTCTCA	AGGATCTTAC	6000
	CGCTGTTGAG	ATCCAGTTCG	ATGTAACCCA	CTCGTGCACC	CAACTGATCT	TCAGCATCTT	6060

TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AACAGGAAG GCAAAATGCC GCAAAAAAGG 6120  
GAATAAGGGC GACACGGAAA TGTGAATAC TCATACTCTT CCTTTTCAA TATTATTGAA 6180  
GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA 6240  
AACAAATAGG GGTTCGCGC ACATTCCCC GAAAAGTGCC ACCT 6284

5

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly  
1 5 10 15  
20 Val His Ser Glu Val Gln Leu Leu Glu Val  
20 25

(2) INFORMATION FOR SEQ ID NO:16:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

35

Ser Pro Gly Lys

1

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 25 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## 10 (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

15 Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly  
 1 5 10 15  
 Val His Ser Glu Leu Thr Gln Ser Pro  
 20 25

## (2) INFORMATION FOR SEQ ID NO:18:

20

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5681 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

30 GACGTCGCGG CCGCTCTAGG CCTCCAAAAA AGCCTCCTCA CTACTTCTGG AATAGCTCAG 60  
 AGGCCGAGGC GGCCTCGGCC TCTGCATAAA TAAAAAAAT TAGTCAGCCA TGCATGGGGC 120  
 GGAGAATGGG CGGAAGTGGG CGGAGTTAGG GCGGGGATGG GCGGAGTTAG GGGCGGGACT 180  
 ATGTTGCTG ACTAATTGAG ATGCATGCTT TGCATACTTC TGCCTGCTGG GGAGCCTGGG 240  
 GACTTTCCAC ACCTGGTTGC TGACTAATTG AGATGCATGC TTTGCATACT TCTGCCTGCT 300  
 35 GGGGAGCCTG GGGACTTTCC ACACCCTAAC TGACACACAT TCCACAGAAT TAATTCCCGG 360  
 GGATCGATCC GTCGACGTAC GACTAGTTAT TAATAGTAAT CAATTACGGG GTCATTAGTT 420



	CATAGCCCAT	ATATGGAGTT	CCGCGTTACA	TAACTTACGG	TAAATGGCCC	GCCTGGCTGA	480
	CCGCCCAACG	ACCCCCGCCC	ATTGACGTCA	ATAATGACGT	ATGTTCCCAT	AGTAACGCCA	540
	ATAGGGACTT	TCCATTGACG	TCAATGGGTG	GACTATTTAC	GGTAAACTGC	CCACTTGCCA	600
	GTACATCAAG	TGTATCATAT	GCCAAGTACG	CCCCCTATTG	ACGTCAATGA	CGGTAAATGG	660
5	CCCGCCTGGC	ATTATGCCCC	GTACATGACC	TTATGGGACT	TTCTTACTTG	GCAGTACATC	720
	TACGTATTAG	TCATCGCTAT	TACCATGGTG	ATGCGGTTTT	GGCAGTACAT	CAATGGGCGT	780
	GGATAGCGGT	TTGACTCACG	GGGATTTCCA	AGTCTCCACC	CCATTGACGT	CAATGGGAGT	840
	TTGTTTTGGC	ACCAAAATCA	ACGGGACTTT	CCAAAATGTC	GTAACAACTC	CGCCCCATTG	900
	ACGCAAATGG	GCGGTAGGCG	TGTACGGTGG	GAGGTCTATA	TAAGCAGAGC	TGGGTACGTG	960
10	AACCGTCAGA	TCGCCTGGAG	ACGCCATCGA	ATTCTGAGCA	CACAGGACCT	CACCATGGGA	1020
	TGGAGCTGTA	TCATCCTCTT	CTTGGTAGCA	ACAGCTACAG	GTGTCCACTC	CGAGCTCACC	1080
	CAGTCTCCAT	CCTCCCTGTC	TGCATCTGTA	GGAGACAGAG	TCACCATCAC	TTGCCGGGCA	1140
	ACTCAGAGTG	TTAGTAACTT	TTTAAATTGG	TATCAGCAGA	AGCCAGGGGA	AGCCCCTACG	1200
	CTCCTGATCT	ATGATGCATC	CACTTCGCAA	AGTGGGGTCC	CATCAAGGTT	CAGTGGCAGT	1260
15	GGATCTGGGA	TGGATTTTCA	TCTCACCATC	AGCAGTCTGC	AGCCTGAAGA	TCTTGCAATG	1320
	TATTACTGTC	AAGCGAGTAT	CAATACCCCG	CTTTTCGGCG	GAGGGACCAG	AATAGATATG	1380
	AGACGAACTG	TGGCTGCACC	ATCTGTCTTC	ATCTTCCCGC	CATCTGATGA	GCAGTTGAAA	1440
	TCTGGAAGTG	CCTCTGTTGT	GTGCCTGCTG	AATAACTTCT	ATCCCAGAGA	GGCCAAAGTA	1500
	CAGTGGAAGG	TGGATAACGC	CCTCCAATCG	GGTAACTCCC	AGGAGAGTGT	CACAGAGCAG	1560
20	GACAGCAAGG	ACAGCACCTA	CAGCCTCAGC	AGCACCCTGA	CGCTGAGCAA	AGCAGACTAC	1620
	GAGAAACACA	AAGTCTACGC	CTGCGAAGTC	ACCCATCAGG	GCCTGAGCTT	GCCCCGCACA	1680
	AAGAGCTTCA	ACAGGGGAGA	GTGTTAGTGA	GATGATCCTC	TAGAGTCATC	TACGTATGAT	1740
	CAGCCTCGAC	TGTGCCTTCT	AGTTGCCAGC	CATCTGTTGT	TTGCCCCCTC	CCCGTGCCTT	1800
	CCTTGACCCT	GGAAGGTGCC	ACTCCCACTG	TCCTTTTCTA	ATAAAATGAG	GAAATTGCAT	1860
25	CGCATTGTCT	GAGTAGGTGT	CATTCTATTC	TGGGGGGTGG	GGTGGGGCAG	GACAGCAAGG	1920
	GGGAGGATTG	GGAAGACAAT	AGCAGGCATG	CTGGGGATGC	GGTGGGCTCT	ATGGAACCAG	1980
	CTGGGGCTCG	ACAGCTCGAG	CTAGCTTTGC	TTCTCAATTT	CTTATTTGCA	TAATGAGAAA	2040
	AAAAGGAAAA	TTAATTTTAA	CACCAATTCA	GTAGTTGATT	GAGCAAATGC	GTTGCCAAAA	2100
	AGGATGCTTT	AGAGACAGTG	TTCTCTGCAC	AGATAAGGAC	AAACATTATT	CAGAGGGAGT	2160
30	ACCCAGAGCT	GAGACTCCTA	AGCCAGTGAG	TGGCACAGCA	TTCTAGGGAG	AAATATGCTT	2220
	GTCATCACCG	AAGCCTGATT	CCGTAGAGCC	ACACCTTGGT	AAGGGCCAAT	CTGCTCACAC	2280
	AGGATAGAGA	GGGCAGGAGC	CAGGGCAGAG	CATATAAGGT	GAGGTAGGAT	CAGTTGCTCC	2340
	TCACATTTGC	TTCTGACATA	GTTGTGTTGG	GAGCTTGGAT	CGATCCACCA	TGGTTGAACA	2400
	AGATGGATTG	CACGCAGGTT	CTCCGGCCGC	TTGGGTGGAG	AGGCTATTTC	GCTATGACTG	2460
35	GGCACAACAG	ACAATCGGCT	GCTCTGATGC	CGCCGTGTTC	CGGCTGTCAG	CGCAGGGGCG	2520
	CCCGGTTCTT	TTTGTCAAGA	CCGACCTGTC	CGGTGCCCTG	AATGAAGTGC	AGGACGAGGC	2580

	AGCGCGGCTA	TCGTGGCTGG	CCACGACGGG	CGTTCCTTGC	GCAGCTGTGC	TCGACGTTGT	2640
	CACTGAAGCG	GGAAGGGACT	GGCTGCATTT	GGGCGAAGTG	CCGGGGCAGG	ATCTCCTGTC	2700
	ATCTCACCTT	GCTCCTGCCG	AGAAAGTATC	CATCATGGCT	GATGCAATGC	GGCGGCTGCA	2760
	TACGCTTGAT	CCGGCTACCT	GCCCATTGCA	CCACCAAGCG	AAACATCGCA	TCGAGCGAGC	2820
5	ACGTACTCGG	ATGGAAGCCG	GTCTTGTGCA	TCAGGATGAT	CTGGACGAAG	AGCATCAGGG	2880
	GCTCGCGCCA	GCCGAAGTGT	TCGCCAGGCT	CAAGGCGCGC	ATGCCCCGACG	GCGAGGATCT	2940
	CGTCGTGACC	CATGGCGATG	CCTGCTTGCC	GAATATCATG	GTGGAAAATG	GCCGCTTTTC	3000
	TGGATTTCATC	GACTGTGGCC	GGCTGGGTGT	GGCGGACCGC	TATCAGGACA	TAGCGTTGGC	3060
	TACCCGTGAT	ATTGCTGAAG	AGCTTGGCGG	CGAATGGGCT	GACCGCTTCC	TCGTGCTTTA	3120
10	CGGTATCGCC	GCTCCCGATT	CGCAGCGCAT	CGCCTTCTAT	CGCCTTCTTG	ACGAGTTCTT	3180
	CTGAGCGGGA	CTCTGGGGTT	CGAAATGACC	GACCAAGCGA	CGCCCAACCT	GCCATCACGA	3240
	GATTTGCGATT	CCACCGCCGC	CTTCTATGAA	AGGTGGGGCT	TCGGAATCGT	TTTCCGGGAC	3300
	GCCGGCTGGA	TGATCCTCCA	GCGCGGGGAT	CTCATGCTGG	AGTTCTTCGC	CCACCCCAAC	3360
	TTGTTTATTG	CAGCTTATAA	TGTTTACAAA	TAAAGCAATA	GCATCACAAA	TTTCACAAAT	3420
15	AAAGCATTTT	TTTCACTGCA	TTCTAGTTGT	GGTTTGTCCA	AACTCATCAA	TGTATCTTAT	3480
	CATGTCTGGA	TCGCGGCCGC	GATCCCGTCG	AGAGCTTGGC	GTAATCATGG	TCATAGCTGT	3540
	TTCTGTGTG	AAATTGTTAT	CCGCTCACAA	TTCCACACAA	CATACGAGCC	GGAAGCATAA	3600
	AGTGTAAGC	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC	ATTAATTGCG	TTGCGCTCAC	3660
	TGCCCCGCTTT	CCAGTCGGGA	AACCTGTCGT	GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	3720
20	CGGGGAGAGG	CGGTTTGCGT	ATTGGGCGCT	CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	3780
	GCTCGGTGCT	TCGGCTGCGG	CGAGCGGTAT	CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	3840
	CCACAGAATC	AGGGGATAAC	GCAGGAAAGA	ACATGTGAGC	AAAAGGCCAG	CAAAAGGCCA	3900
	GGAACCGTAA	AAAGGCCGCG	TTGCTGGCGT	TTTTCCATAG	GCTCCGCCCC	CCTGACGAGC	3960
	ATCACAAAAA	TCGACGCTCA	AGTCAGAGGT	GGCGAAACCC	GACAGGACTA	TAAAGATACC	4020
25	AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC	GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	4080
	GATACCTGTC	CGCCTTTCTC	CCTTCGGGAA	GCGTGGCGCT	TTCTCAATGC	TCACGCTGTA	4140
	GGTATCTCAG	TTGCGGTGTAG	GTCGTTGCGT	CCAAGCTGGG	CTGTGTGCAC	GAACCCCCCG	4200
	TTCAGCCCCG	CCGCTGCGCC	TTATCCGGTA	ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	4260
	ACGACTTATC	GCCACTGGCA	GCAGCCACTG	GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	4320
30	GCGGTGCTAC	AGAGTTCTTG	AAGTGGTGGC	CTAACTACGG	CTACACTAGA	AGGACAGTAT	4380
	TTGGTATCTG	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	4440
	CCGGCAAACA	AACCACCGCT	GGTAGCGGTG	GTTTTTTTGT	TTGCAAGCAG	CAGATTACGC	4500
	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	TACGGGGTCT	GACGCTCACT	4560
	GGAACGAAAA	CTCACGTTAA	GGGATTTTGG	TCATGAGATT	ATCAAAAAGG	ATCTTCACCT	4620
35	AGATCCCTTT	AAATTAAAAA	TGAAGTTTAA	AATCAATCTA	AAGTATATAT	GAGTAAACTT	4680
	GGTCTGACAG	TTACCAATGC	TTAATCAGTG	AGGCACCTAT	CTCAGCGATC	TGTCTATTTC	4740

GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG GAGGGCTTAC 4800  
 CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT 4860  
 CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA ACTTTATCCG 4920  
 CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA 4980  
 5 GTTTGCGCAA CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGTA 5040  
 TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC CCCATGTTGT 5100  
 GCAAAAAAGC GGTAGCTCC TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGAG 5160  
 TGTTCATCACT CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAA 5220  
 GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC 5280  
 10 GACCGAGTTG CTCTTGCCCG GCGTAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT 5340  
 TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG ATCTTACCGC 5400  
 TGTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA 5460  
 CTTTCACCAG CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA 5520  
 TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA 5580  
 15 TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC 5640  
 AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC T 5681

## (2) INFORMATION FOR SEQ ID NO:19:

## 20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

30 Leu Pro Val Thr Lys Ser Phe Asn Arg Gly Glu Cys  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO:20:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1427 base pairs

- (B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAATTCGGTA CCATGGAGTT TGGGCTGAGC TGGGTTTTC TCGTGGCTCT TTTAAGAGGT 60  
GTCCAGTGTC AGGTGCAGCT GGTGGAGTCT GGGGGAGGCC TGGTCAGGCC TGGCGGGTCC 120  
10 CTAAGACTCT CGTGTGCAGC CTCTGGAACC ACCCTCAGTG GCTATACCAT GCACTGGGTC 180  
CGCCAGGCTC CAGGGAAGGG GCTGGAGTGG GTCTCATCCA TTACTGGAGG TAGCAACTTC 240  
ATAAACTACT CAGACTCAGT GAAGGGCCGA TTCACCATCT CCAGAGACAA CGCCAAGAAC 300  
TCACTTTATC TGCAAATGAA CAGCCTGACA GCCGAGGACA CGGCTGTCTA TTATTGTGCG 360  
ACCGCCCCTA TAGCACCGCC CTACTTTGAC CACTGGGGCC AGGGAACCCT GGTCACCGTC 420  
15 TCCTCAGCCT CCACCAAGGG CCCATCGGTC TTCCCCCTGG CACCCTCCTC CAAGAGCACC 480  
TCTGGGGGCA CAGCGGCCCT GGGCTGCCTG GTCAAGGACT ACTTCCCCGA ACCGGTGACC 540  
GTGTCGTGGA ACTCAGGCGC CCTGACCAGC GGCCTGCACA CCTTCCCGGC TGTCTACAG 600  
TCCTCAGGAC TCTACTCCCT CAGCAGCGTG GTGACCGTGC CCTCCAGCAG CTTGGGCACC 660  
CAGACCTACA TCTGCAACGT GAATCACAAG CCCAGCAACA CCAAGGTGGA CAAGAAAGTT 720  
20 GAGCCCAAAT CTTGTGACAA AACTCACACA TGCCACCGT GCCCAGCACC TGAACCTCTG 780  
GGGGGACCGT CAGTCTTCCT CTTCCCCCA AAACCAAGG ACACCCTCAT GATCTCCCGG 840  
ACCCCTGAGG TCACATGCGT GGTGGTGGAC GTGAGCCACG AAGACCCTGA GGTCAAGTTC 900  
AACTGGTACG TGGACGGCGT GGAGGTGCAT AATGCCAAGA CAAAGCCGCG GGAGGAGCAG 960  
TACAACAGCA CGTACCGGGT GGTACGCGTC CTCACCGTCC TGCACCAGGA CTGGCTGAAT 1020  
25 GGCAAGGAGT ACAAGTGCAA GGTCTCCAC AAAGCCCTCC CAGCCCCAT CGAGAAAACC 1080  
ATCTCCAAAG CCAAAGGGCA GCGCCGAGAA CCACAGGTGT ACACCCTGCC CCCATCCCGG 1140  
GATGAGCTGA CCAAGAACCA GGTACGCTG ACCTGCCTGG TCAAAGGCTT CTATCCAGC 1200  
GACATCGCCG TGGAGTGGGA GAGCAATGGG CAGCCGAGA ACAACTACAA GACCACGCCT 1260  
CCCGTGCTGG ACTCCGACGG CTCCTTCTTC CTCTACAGCA AGCTCACCGT GGACAAGAGC 1320  
30 AGGTGGCAGC AGGGGAACGT CTTCTCATGC TCCGTGATGC ATGAGGCTCT GCACAACCAC 1380  
TACACGCAGA AGAGCCTCTC CCTGTCTCCG GGTAAATGAT AGATATC 1427

(2) INFORMATION FOR SEQ ID NO:21:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

- (B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Leu Leu Arg Gly  
10 1 5 10 15  
Val Gln Cys Gln Val Gln Leu Val  
20

(2) INFORMATION FOR SEQ ID NO:22:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 732 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

25 GAATTCATG GACATGAGGG TCCCCGCTCA GCTCCTAGGG CTCCTGCTGC TCTGGCTCCG 60  
AGGTGCCAGA TGTGACATCC AGATGACCCA CTCTCCATCC TCCCTGTCTG CATCTGTAGG 120  
AGACAGAGTC ACCATCACTT GCCGGGCAAC TCAGAGTGTT AGTAACTTTT TAAATTGGTA 180  
TCAGCAGAAG CCAGGGGAAG CCCCTACGCT CCTGATCTAT GATGCATCCA CTTGCGAAAG 240  
TGGGGTCCCA TCAAGGTTCA GTGGCAGTGG ATCTGGGATG GATTTTCAGTC TCACCATCAG 300  
30 CAGTCTGCAG CCTGAAGATC TTGCAATGTA TTAAGTGTCAA GCGAGTATCA ATACCCCGCT 360  
TTTCGGCGGA GGGACCAGAA TAGATATGAG ACGAACTGTG GCTGCACCAT CTGTCTTCAT 420  
CTTCCCGCCA TCTGATGAGC AGTTGAAAATC TGGAACTGCC TCTGTTGTGT GCCTGCTGAA 480  
TAACTTCTAT CCCAGAGAGG CCAAAGTACA GTGGAAGGTG GATAACGCCC TCCAATCGGG 540  
TAACTCCCAG GAGAGTGTCA CAGAGCAGGA CAGCAAGGAC AGCACCTACA GCCTCAGCAG 600  
35 CACCCTGACG CTGAGCAAAG CAGACTACGA GAAACACAAA GTCTACGCCT GCGAAGTCAC 660  
CCATCAGGGC CTGAGCTTGC CCGTCACAAA GAGCTTCAAC AGGGGAGAGT GTTAGTGAGA 720

TGATCCTCTA GA

732

## (2) INFORMATION FOR SEQ ID NO:23:

## 5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

15 Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp Leu Arg  
1 5 10 15  
Gly Ala Arg Cys Asp Ile Gln Met Thr  
20 25

## 20 (2) INFORMATION FOR SEQ ID NO:24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- 25 (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Leu Pro Val Thr Lys Ser Phe Asn Arg Gly Glu Cys  
1 5 10

## 35 (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1427 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

10 GAATTCGGTA CCATGGAGTT TGGGCTGAGC TGGGTTTTTC TCGTGGCTCT TTAAAGAGGT 60  
 GTCCAGTGTC AGGTGCAGCT GGTGGAGTCT GGGGGAGGCC TGGTCAGGCC TGGCGGGTCC 120  
 CTAAGACTCT CGTGTGCAGC CTCTGGAACC ACCCTCAGTG GCTATAACCAT GCACTGGGTC 180  
 CGCCAGGCTC CAGGGAAGGG GCTGGAGTGG GTCTCATCCA TTAAGGAGG TAGCAACTTC 240  
 ATAAACTACG CAGACTCAGT GAAGGGCCGA TTCACCATCT CCAGAGACAA CGCCAAGAAC 300  
 15 TCACTTTATC TGCAAATGAA CAGCCTGACA GCCGAGGACA CGGCTGTCTA TTATTGTGCG 360  
 ACCGCCCCCTA TAGCACCGCC CTACTTTGAC CACTGGGGCC AGGGAACCCCT GGTCACCGTC 420  
 TCCTCAGCCT CCACCAAGGG CCCATCGGTC TTCCCCCTGG CACCCTCCTC CAAGAGCACC 480  
 TCTGGGGGCA CAGCGGCCCT GGGCTGCCTG GTCAAGGACT ACTTCCCCGA ACCGGTGACC 540  
 GTGTCGTGGA ACTCAGGCGC CCTGACCAGC GCGGTGCACA CCTTCCCGGC TGTCTTACAG 600  
 20 TCCTCAGGAC TCTACTCCCT CAGCAGCGTG GTGACCGTGC CCTCCAGCAG CTGGGGCACC 660  
 CAGACCTACA TCTGCAACGT GAATCACAAG CCCAGCAACA CCAAGGTGGA CAAGAAAGTT 720  
 GAGCCCAAAT CTTGTGACAA AACTCACACA TGCCCACCGT GCCCAGCACC TGAAGTCTCTG 780  
 GGGGGACCGT CAGTCTTCCT CTCCCCCA AAACCCAAGG ACACCCTCAT GATCTCCCGG 840  
 ACCCCTGAGG TCACATGCGT GGTGGTGGAC GTGAGCCACG AAGACCCTGA GGTCAAGTTC 900  
 25 AACTGGTACG TGGACGGCGT GGAGGTGCAT AATGCCAAGA CAAAGCCGCG GGAGGAGCAG 960  
 TACAACAGCA CGTACCGGT GGTGAGCGTC CTCACCGTCC TGCACCAGGA CTGGCTGAAT 1020  
 GGCAAGGAGT ACAAGTGCAA GGTCTCCAAC AAAGCCCTCC CAGCCCCCAT CGAGAAAACC 1080  
 ATCTCCAAAG CCAAAGGGCA GCGCCGAGAA CCACAGGTGT ACACCCTGCC CCCATCCCGG 1140  
 GATGAGCTGA CCAAGAACCA GGTGAGCCTG ACCTGCCTGG TCAAAGGCTT CTATCCAGC 1200  
 30 GACATCGCCG TGGAGTGGGA GAGCAATGGG CAGCCGGAGA ACAACTACAA GACCACGCCT 1260  
 CCCGTGCTGG ACTCCGACGG CTCCTTCTTC CTCTACAGCA AGCTCACCGT GGACAAGAGC 1320  
 AGGTGGCAGC AGGGGAACGT CTTCTCATGC TCCGTGATGC ATGAGGCTCT GCACAACCAC 1380  
 TACACGCAGA AGAGCCTCTC CCTGTCTCCG GGTAAATGAT AGATATC 1427

35

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Ser Asn Phe Ile Asn Tyr Ala

1

5

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1427 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

25 GAATTCGGTA CCATGGAGTT TGGGCTGAGC TGGGTTTTCC TCGTGGCTCT TTAAAGAGGT 60  
GTCCAGTGTC AGGTGCAGCT GGTGGAGTCT GGGGGAGGCC TGGTCAGGCC TGGCGGGTCC 120  
CTAAGACTCT CGTGTGCAGC CTCTGGAACC ACCCTCAGTG GCTATACCAT GCACTGGGTC 180  
CGCCAGGCTC CAGGGAAGGG GCTGGAGTGG GTCTCATCCA TTAAGGAGG TAGCAACTTC 240  
ATACAATACT CAGACTCAGT GAAGGGCCGA TTCACCATCT CCAGAGACAA CGCCAAGAAC 300  
30 TCACCTTATC TGCAAATGAA CAGCCTGACA GCCGAGGACA CGGCTGTCTA TTATTGTGCG 360  
ACCGCCCTA TAGCACC GCC CTACTTTGAC CACTGGGGCC AGGGAACCCT GGTACCCGTC 420  
TCCTCAGCCT CCACCAAGGG CCCATCGGTC TTCCCCTGG CACCCTCCTC CAAGAGCACC 480  
TCTGGGGGCA CAGCGGCCCT GGGCTGCCTG GTCAAGGACT ACTTCCCCGA ACCGGTGACC 540  
GTGTCGTGGA ACTCAGGCGC CCTGACCAGC GGCCTGCACA CCTTCCCGGC TGTCTACAG 600  
35 TCCTCAGGAC TCTACTCCCT CAGCAGCGTG GTGACCGTGC CCTCCAGCAG CTTGGGCACC 660  
CAGACCTACA TCTGCAACGT GAATCACAAG CCCAGCAACA CCAAGGTGGA CAAGAAAGTT 720



5 GAGCCCAAAT CTTGTGACAA AACTCACACA TGCCACCGT GCCCAGCACC TGAACCTCTG 780  
GGGGGACCGT CAGTCTTCCT CTTCCCCCA AAACCAAGG ACACCCTCAT GATCTCCCGG 840  
ACCCCTGAGG TCACATGCGT GGTGGTGGAC GTGAGCCACG AAGACCCTGA GGTCAAGTTC 900  
AACTGGTACG TGGACGGCGT GGAGGTGCAT AATGCCAAGA CAAAGCCGCG GGAGGAGCAG 960  
TACAACAGCA CGTACCGGGT GGTGAGCGTC CTCACCGTCC TGCACCAGGA CTGGCTGAAT 1020  
GGCAAGGAGT ACAAGTGCAA GGTCTCCAAC AAAGCCCTCC CAGCCCCCAT CGAGAAAACC 1080  
ATCTCCAAAG CCAAAGGGCA GCCCCGAGAA CCACAGGTGT ACACCCTGCC CCCATCCCGG 1140  
GATGAGCTGA CCAAGAACCA GGTGAGCCTG ACCTGCCTGG TCAAAGGCTT CTATCCCAGC 1200  
GACATCGCCG TGGAGTGGGA GAGCAATGGG CAGCCGAGA ACAACTACAA GACCACGCCT 1260  
10 CCCGTGCTGG ACTCCGACGG CTCCTTCTTC CTCTACAGCA AGCTCACCGT GGACAAGAGC 1320  
AGGTGGCAGC AGGGGAACGT CTTCTCATGC TCCGTGATGC ATGAGGCTCT GCACAACCAC 1380  
TACACGCAGA AGAGCCTCTC CCTGTCTCCG GGTAAATGAT AGATATC 1427

## (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Ser Asn Phe Ile Gln Tyr Ser  
1 5

## (2) INFORMATION FOR SEQ ID NO:29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 762 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAATTCATG GACATGAGGG TCCCCGCTCA GCTCCTAGGG CTCCTGCTGC TCTGGCTCCG 60  
 5 AGGTGCCAGA TGTGACATCC AGATGACCCA GTCTCCATCC TCCCTGTCTG CATCTGTAGG 120  
 AGACAGAGTC ACCATCACTT GCCGGGCAAC TCAGAGTGTT AGTAACTTTT TAAATTGGTA 180  
 TCAGCAGAAG CCAGGGGAAG CCCCTACGCT CCTGATCTAT GATGCATCCA CTTGCGAAAG 240  
 TGGGGTCCCA TCAAGGTTCA GTGGCAGTGG ATCTGGGATG GATTTCAGTC TCACCATCAG 300  
 CAGTCTGCAG CCTGAAGATC TTGCAATGTA TTA CTGTCAA GCGAGTATCA ATACCCCGCT 360  
 10 TTTCGGCGGA GGGACCAGAA TAGATATGAG ACGAACTGTG GCTGCACCAT CTGTCTTCAT 420  
 CTTCCCGCCA TCTGATGAGC AGTTGAAATC TGGAAGTGGC TCTGTTGTGT GCCTGCTGAA 480  
 TAACTTCTAT CCCAGAGAGG CCAAAGTACA GTGGAAGGTG GATAACGCCC TCCAATCGGG 540  
 TAACTCCCAG GAGAGTGTC CAGAGCAGGA CAGCAAGGAC AGCACCTACA GCCTCAGCAG 600  
 CACCCTGACG CTGAGCAAAG CAGACTACGA GAAACACAAA GTCTACGCCT GCGAAGTCAC 660  
 15 CCATCAGGGC CTGAGCTCGC CCGTCACAAA GAGCTTCAAC AGGGGAGAGT GTTAGTGAGA 720  
 TGATCCTCTA GATCTACGTA TGATCAGCCT CGACTGTGCC TT 762

## (2) INFORMATION FOR SEQ ID NO:30:

- 20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 12 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- 25 (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

30 Ser Pro Val Thr Lys Ser Phe Thr Arg Gly Gln Cys  
 1 5 10